U.S. Research Vessel Surface Meteorology Data Assembly Center

Shawn R. Smith, Mark A. Bourassa and Eric Chassignet

Center for Ocean-Atmospheric Prediction Studies, The Florida State University, Tallahassee FL

1. PROJECT SUMMARY

The central activity of the U.S. Research Vessel Surface Meteorology Data Assembly Center (DAC) is the continued development of the Shipboard Automated Meteorological and Oceanographic System (SAMOS) initiative (http://samos.coaps.fsu.edu/). The SAMOS initiative focuses on improving the quality of and access to surface marine meteorological and oceanographic data collected *in-situ* by automated instrumentation on research vessels and ships of opportunity. The DAC activities focus primarily on NOAA Strategic Plan Goals 2 and 3 by providing high quality weather and near surface ocean data for use in validating satellite products, global air-sea flux analyses, and model fields. Research vessels are mobile observing platforms that are an essential component of the global ocean observing system. These vessels travel to remote and hard to observe ocean locations that are far from normal shipping lanes.

The rationale for this activity centers on the desire to understand the physical and thermodynamic interaction between the ocean and atmosphere. This interaction is key to our understanding of how marine weather systems evolve, how these systems impact the ocean, and how the oceans impact the weather. On longer time scales, understanding the interaction between the ocean and atmosphere is necessary to assess our changing global climate system. The role of the DAC is providing the high quality marine meteorological and surface ocean measurements to the research and operational community so that they can address these ocean-atmospheric interactions. High quality observations are essential to our scientific understanding of the ocean-atmosphere interactions.

The DAC was established at the Florida State University specifically to coordinate the collection, quality evaluation, distribution, and future archival of SAMOS data. SAMOS are typically a computerized data logging system that continuously records navigation (ship's position, course, speed, and heading), meteorological (winds, air temperature, pressure, moisture, rainfall, and radiation), and near ocean surface (sea temperature and salinity) parameters while a vessel is at sea. Measurements are recorded at high-temporal sampling rates (typically 1 minute or less). The DAC collaborated with the Woods Hole Oceanographic Institution (WHOI) to design a ship-to-shore-to-user data pathway for U.S. research vessel SAMOS data. In the past, the data flowed from ship to shore only in a delayed-mode with a 3 month to 2-year lag between collection and availability to the user community. The new data pathway supports automated data transmission from each ship to the DAC on a daily basis. A "preliminary" version of the SAMOS data are available on-line within 5 minutes of receipt by the DAC. The preliminary data undergo common formatting, metadata enhancement, and automated quality control. Visual inspection and further scientific quality control result in a "research" quality SAMOS product that are nominally distributed with a delay of 10 days from the original data collection date. All quality-evaluated research vessel data from the SAMOS initiative and past DAC programs are freely available to the user community (http://www.coaps.fsu.edu/RVSMDC/html/data.shtml), and we continue to

work with several world data center archives (e.g., National Oceanographic Data Center, National Center for Atmospheric Research) to ensure long term stewardship of these data.

2. ACCOMPLISHMENTS

Over the past year, our efforts have focused on the continued development of the SAMOS Initiative. In collaboration with partners at NOAA's Office of Marine and Aviation Operations (OMAO), Raytheon Polar Services (RPS), and the United States Coast Guard (USCG), the number of vessels participating in SAMOS has increased from 12 to 17. We continue to actively recruit new vessels to participate in the SAMOS initiative. Data from all 17 vessels are routinely pushed through our automated (preliminary) data quality evaluation (DQE). In the past year our research quality processing (including visual DQE) has been operationally applied, but a backlog of SAMOS data still needs to complete the research processing. Many upgrades have been made to both our public access web site and our internal data base tools. Throughout the year, DAC personnel have been actively promoting the SAMOS Initiative at both national and international meetings.

Deliverables for FY 2008 included:
1. Continue routine quality evaluation of meteorological data for 12 SAMOS vessels currently contributing to the DAC
2. Recruit additional research vessels to the SAMOS initiative (focus will be on NOAA and UNOLS vessels)
3. Improve and enhance metadata collection from participating research vessels
4. Implement additional data quality evaluation techniques.
5. Create and distribute turbulent air-sea fluxes for vessels contributing to SAMOS
    - Subtask1: Compare SAMOS fluxes with available NWP products
6. Continue liaison activities with U.S. and international government agencies, archives, climate programs
    - Subtask1: Establish data exchange with GOSUD
    - Subtask2: Implement data exchange with NODC and NCAR archives
    - Subtask3: Produce subset of SAMOS observations for ICOADS
7. Expand user options on the SAMOS web site to ease user access to observations and metadata

The following accomplishments address the deliverables. Also noted are impediments to achieving the deliverables.

2.1. Vessel recruitment [Deliverable 2]

Recruitment of additional vessels to participate in the SAMOS Initiative was very successful during the reporting period. Five new vessels were recruited (Table 1) in the past year. These vessels now routinely contribute SAMOS observations when they are at sea. Collaboration with NOAA OMAO resulted in four new recruitments. The *Oceanus*, operated by WHOI, is the first UNOLS vessel to be recruited since 1995. We also have been in communication with Eric Schulz from the Australian Bureau of Meteorology to recruit two Australian RVs. A data

exchange framework is in place that will involve the Australian Bureau sending their meteorological data to the DAC in SAMOS format netCDF files. We will conduct automated QC of these files and will distribute them with other SAMOS data. Visual QC will only be done in Australia and we are working to establish a protocol to distribute their visually inspected data.

In addition, Co-PI Smith and Jeremy Rolph (our data quality analyst) attended the annual UNOLS RVTEC meeting November 2007 and participated in the 2nd Joint GOSUD/SAMOS workshop. These (and other) workshops provided and opportunity to recruit additional vessel operators. Most expressed interest in participating in SAMOS and have made verbal commitments, but initiating new data transfers is still difficult in these times of tight operational budgets for research vessels. Flat funding at the SAMOS DAC has also limited our ability to further increase the number of vessels contributing data to the DAC.

Table 1. Ships transmitting observations to SAMOS DAC during FY 2007 and FY 2008.

		Number of ship days with data	
Vessel	Operator	1/10/2006 – 30/9/2007	1/10/2007 – 30/9/2008
Atlantis	WHOI	291	294
David Star Jordan	NOAA		88
Fairweather	NOAA		79
Gordon Gunter	NOAA	4	44
Healy	USCG	62	162
Henry Bigelow	NOAA	70	139
Hi'ialakai	NOAA	107	196
Ka'imimoana	NOAA	61	197
Knorr	WHOI	252	305
Lawrence Gould	NSF/Raytheon	32	260
Miller Freeman	NOAA	190	187
Nancy Foster	NOAA	122	179
Oceanus	WHOI		115
Oregon	NOAA		84
Oscar Dyson	NOAA	170	211
Rainier	NOAA		107
Ronald Brown	NOAA	149	151
		1510	2798

2.2. Daily SAMOS data processing [Deliverable 1]

Preliminary processing of SAMOS observations received via daily email messages from participating research vessels is an ongoing operational activity at the DAC. During FY 2008, 2798 days of shipboard meteorology data were processed for the 17 recruited vessels (see Table 1). Preliminary processing (Figure 1) starts once the data file arrives at the DAC as an attachment to an email. Each email attachment is unpacked, the data provided are verified that they conform to the format and parameters expected for the individual vessel, and finally the data are blended with vessel specific metadata and are converted to a common netCDF format. The data for each day are then passed through an automated quality evaluation program and data quality statistics are calculated prior to the file being posted for users on the SAMOS web and ftp

sites. The entire process from arrival at the DAC to distribution of the preliminary data files is fully automated. Preliminary files appear on the data distribution site within 5 minutes of their arrival at the DAC (typically shortly after 0000 UTC).

A comparison of the spatial distribution of data received, processed, and on-line for FY 2007 and FY 2008 is shown in Figure 2. The addition of 5 ships and more complete data from the 12 recruited in the previous year resulted in over 1200 more ship days of data than in FY 2007. We continue to track all stages of the quality processing of the ship data in the SAMOS SQL database and quality statistics are accessible via the SAMOS web site (http://samos.coaps.fsu.edu/html/data_availability.php).

2.3. Delayed-mode SAMOS processing [Deliverables 1 and 4]

Due to data logging problems on the ship or communication dropouts, some data arrive several days after they were collected. Often the data are noted to be missing by the analyst at the DAC and arrive after the analyst notifies the vessel technician at sea. In addition, data for a single day may be fragmented and may arrive in multiple files attached to a single email. As a result, the DAC developed a method to merge multiple files for a single observing day into a combined, delayed-mode data file. This merged file undergoes additional automated (e.g., duplicate removal) and visual data quality evaluation and is then released as a "research-quality" SAMOS data file for the particular observation day (Figure 1).

Currently the merge occurs 10 calendar days after the observation day (when the preliminary data should arrive at the DAC). Nominally, the data quality analyst at COAPS reviews the latest merged files and conducts a visual quality evaluation on a daily basis. The visual analysis is accomplished using SVIDAT, a graphical user interface developed by COAPS programmers. SVIDAT allows the analyst to review, add, or modify data quality flags on the merged files. Once the analyst is satisfied with the data quality, the file is saved and posted automatically to the SAMOS ftp and web sites. This process also updates all necessary tracking information in the ship database and creates the copies of the original, preliminary, and research quality files for delivery to the national archive centers.

The rapid increase in data volume flowing through the SAMOS DAC (over 1200 more ship days of data in FY 2008; Table 1) has overwhelmed the ability of our visual analyst to keep up with the delayed-mode visual inspection. The analyst has been working forward from September 2007 to complete the delayed-mode inspection and prepare the data for final archival at NODC. The number of ship days of data ready for archival are shown in Table 2. Comparing to Table 1, one can see that there are nearly 2000 more days to evaluate for 2008 (and new data arrives each day). Through separate funding at COAPS, we hired a temporary data analyst to help with this backlog; however, we anticipate that this will be sufficient to clear the backlog. Support for this analyst is temporary and without additional funding, this analyst will not be retained. Designing methods to streamline the QC system to further reduce visual inspection will be a priority for FY 2009. A new statistical technique for delayed-mode QC is currently under development.

2.4. Metadata enhancement [Deliverable 3]

The SAMOS DAC continues to work to improve access to the metadata (e.g., instrument heights, sensor locations, averaging methods, etc.) necessary for scientific application of these observations. This continues to be a struggle. It is easier to get the SAMOS data from recruited vessels, but some critical metadata is lacking. We have implemented a web-based metadata interface that allows the SAMOS data providers to update their metadata via a secure login for each operator and ship. We have provided demonstrations of this interface at RVTEC and the 2nd GOSUD/SAMOS workshop. Feedback from users has still been limited and we will continue to push forward with training on this interface.

Table 2. Data completing research quality processing at the SAMOS DAC and ready for archival at NODC for FY2007 and FY2008.

		Number of ship days ready for archival	
Vessel	Operator	1/10/2006 – 30/9/2007	1/10/2007 – 30/9/2008
Atlantis	WHOI	92	120
David Star Jordan	NOAA		4
Fairweather	NOAA		5
Gordon Gunter	NOAA		
Healy	USCG	30	6
Henry Bigelow	NOAA	9	34
Hi'ialakai	NOAA	23	49
Ka'imimoana	NOAA	11	77
Knorr	WHOI	29	129
Lawrence Gould	NSF/Raytheon		29
Miller Freeman	NOAA	28	71
Nancy Foster	NOAA	21	54
Oceanus	WHOI		
Oregon	NOAA		
Oscar Dyson	NOAA	27	59
Rainier	NOAA		2
Ronald Brown	NOAA	15	90
		285	729

Another new feature is date stamping all metadata within the SAMOS ship profile database. This allows metadata changes to be tracked and will ensure that the proper metadata is assigned to the SAMOS data files (even in the event of reprocessing earlier data to include late arriving reports).

The issue of metadata collection and access is a broad problem. Mr. Smith continues to serve on the Data Management Best Practices Subcommittee of the UNOLS council. The committee is reviewing data management practices (including metadata) for the UNOLS fleet and has developed a framework for required cruise level metadata from each vessel for each cruise. Adoption of this framework is still pending in the UNOLS council, but once it is accepted, the DAC will work with UNOLS to share cruise level metadata for all SAMOS vessels. Metadata related to specific instruments is even more of a challenge; so one outcome of the 2nd Joint GOSUD/SAMOS workshop was to establish a SAMOS task team to define an automated method to transfer instrument metadata from ship to shore on a regular schedule. This task team

will be organized and will begin their work in late 2008. The team will consider the practical needs of vessel operators and the larger marine metadata initiatives of JCOMM, etc.

2.5. Liaison activities [Deliverable 6]

The SAMOS DAC serves as the project office for the entire SAMOS initiative. In this capacity, DAC personnel facilitate U.S. and international collaborations on topics ranging from data accuracy, data acquisition and exchange, training activities, and data archival. Major meetings include Ocean Sciences (March 2008) and CLIMAR-III (May 2008). We have forged new partnerships within the metadata community (NSF Legacy workshop, Sept. 2008), educators of marine technicians through the Marine Advanced Technology Education (MATE) center (MATE partner meeting, August 2008), and the coastal observing community (Northern Gulf Institute Meeting, May 2008).

Foremost of the liaison activities was convening the 2nd Joint GOSUD/SAMOS workshop in Seattle, Washington from 10-12 June 2008. This meeting was supported via additional OCO funding separate from our base support. UCAR JOSS arranged and the United States Coast Guard hosted the meeting. The workshop organizing committee (Shawn Smith, Mark Bourassa, Loic Petit de la Villéon, David Forcucci, and Phillip McGillivary) brought together a panel consisting of operational and research scientists, educators, marine technicians, and private sector and government representatives. Broad topic areas included new opportunities for international collaboration, emerging technologies, scientific application of underway measurements, and data and metadata issues. New sessions included a technician's round-table discussion and developing educational initiatives. Scientific discussion centered around the need for high-quality meteorological and thermosalinograph observations to support satellite calibration and validation, ocean data assimilation, polar studies, air-sea flux estimation, and improving analyses of precipitation, carbon, and radiation. Determining the regions of the ocean and observational parameters necessary to achieve operational and research objectives requires input by the scientific user community (e.g., via CLIVAR). This input will allow SAMOS and GOSUD to target their limited resources on vessels operating in the high priority regions. The vessel operators and marine technicians were very supportive of the activities of SAMOS and GOSUD. They requested a clear set of guidelines for parameters to measure, routine monitoring activities, and calibration schedules. The operators also desire additional routine feedback on data flow and data quality. A clear need for educational materials was noted by the technical community. The dissemination of best practices guides for existing techs and pre-cruise training for new techs were suggested. The result of the workshop was a series of action items and ten recommendations. The resulting action items will help address subtask 1 (initiating an exchange between GOSUD and SAMOS). The full workshop report, objectives, agenda, poster and oral presentations, and list of participants are available at http://www.coaps.fsu.edu/RVSMDC/marine_workshop4/.

2.6. Data archival [Deliverable 6, subtask 2]

Chris Paver from NODC visited COAPS in late August 2008 to learn about the SAMOS QC procedures and to establish an archival agreement with NODC. The meeting resulted in COAPS and NODC drafting a submission agreement that will guide the routine transmission of SAMOS

data to NODC. Using an rsync protocol, NODC will be able to routinely receive a full set of SAMOS data for each participating vessel. The protocol ensures that a complete archive set for each ship will be provided to NODC once a full month of data has completed research quality QC at COAPS. The full archive set will include:

- the original ASCII formatted data from the vessel (as received via email)
- both preliminary and research quality controlled SAMOS netCDF files
- a file map showing the relationship between the original and netCDF files
- a variable map linking the original variable names (from the provider) to the variable names used in the netCDF file (operationally this is stored in our SQL database)
- documentation (readme, file naming, file format, and processing procedures)
- md5 check sum file

NODC will know the month of data is complete and ready for archival (and assignment of an NODC accession number) once the MD5 check sum file is available. The interval between updates for a given ship will nominally be monthly, once the backlog of research quality data is completed (see above). We are also exploring NODC providing off-site backup of all critical SAMOS codes and data in the event of a catastrophic event (fire, hurricane, etc.).

2.7. Data Distribution [Deliverable 7]

All near real-time (quick) and delayed-mode (research) data are available via the web (http://samos.coaps.fsu.edu/, under "Data Access") and ftp (samos.coaps.fsu.edu, anonymous access, cd /samos_pub/data/) sites. The SAMOS web site also includes an overview of the initiative, provides links to relevant literature and best practices guides, and access to past SAMOS workshops. The DAC provides access to the preliminary quality controlled data for all 17 ships currently recruited to the SAMOS initiative, and access to research quality data from September 2007 – February 2008 is new in FY 2008. A searchable metadata portal allows users to access ship- and parameter-specific metadata. An improved interface for digital photos and schematics of participating vessels is now available which includes pop-up descriptors of each graphic. The web site also provides access to recruitment materials for vessels, desired SAMOS parameters, accuracy requirements, and training materials.

The data distribution system is under constant development. A major new feature developed in FY 2008 is a user interface that allows selection by ship(s) and dates to create cruise track maps on Google Earth (http://samos.coaps.fsu.edu/html/ship_tracker.php). The cruise tracks are plotted by accessing data for each ship that has been sub-sampled at an hourly interval (from the original one-minute interval) and stored in the SAMOS SQL database. Currently the sub-sample is the one-minute value at the top of each hour, but this will be changed to a 10-minute average to support the ICOADS project sometime in FY 2009 (Deliverable 6, subtask 3). The track map provides cruise lines that are automatically color coded by ship. The mapping tool allows zoom and swapping from a satellite to a hybrid map available from Google Earth. Examples of the maps are shown in figure 2.

At present, none of the SAMOS data are distributed in real-time via the Global Telecommunications System (GTS). The primary reasons are based on cost and the desire to ensure that the SAMOS data can fulfill their primary goal as an independent validation source for models and satellite products. Over the past year, there has been growing interest (expressed

at CLIMAR-III and the 2nd Joint GOSUD/SAMOS workshop) to place a subset of the SAMOS observations on the GTS (even in delayed mode). The DAC is exploring this option, but the practicalities of ensuring that a SAMOS report can be distinguished from a regular Voluntary Observing Ship report from the same vessel need to be carefully explored. A modeling center should be able to differentiate a quality-controlled SAMOS report collected using research quality sensors from the manual bridge observation for that vessel (and differentiate the metadata from both). Current GTS formats are not well suited to this task. A dialog has been implemented with the U. S. VOS coordinator, the SOT from JCOMM, and the ICOADS project to determine the best approach to placing SAMOS data on the GTS. A series of "masked" call signs is being considered by SOT for SAMOS (G. Ball, personal communication, 2008). The other reality is that placing SAMOS data on the GTS will require additional programming at the DAC to place data in the desired format (BUFR or SHIP), coordination with JCOMM-OPS to build the masked call sign list, and careful coordination with WMO concerning metadata issues. These all will take additional resources that are not available under our proposed flat budget.

2.8. Deliverable 5 pushed forward to FY 2008

Difficulties in obtaining the necessary metadata to accurately compute the turbulent fluxes and the backlog of data needing research quality QC have slowed progress on this task. We are making progress in both areas and anticipate preliminary flux work to begin in FY 2009.
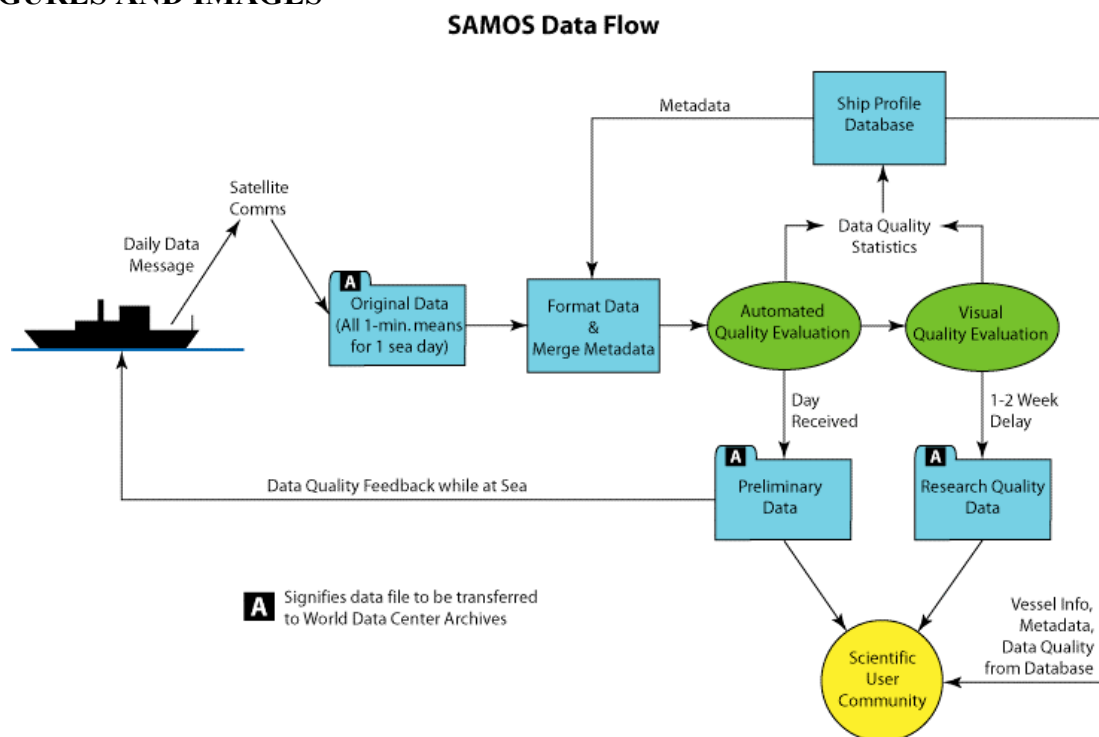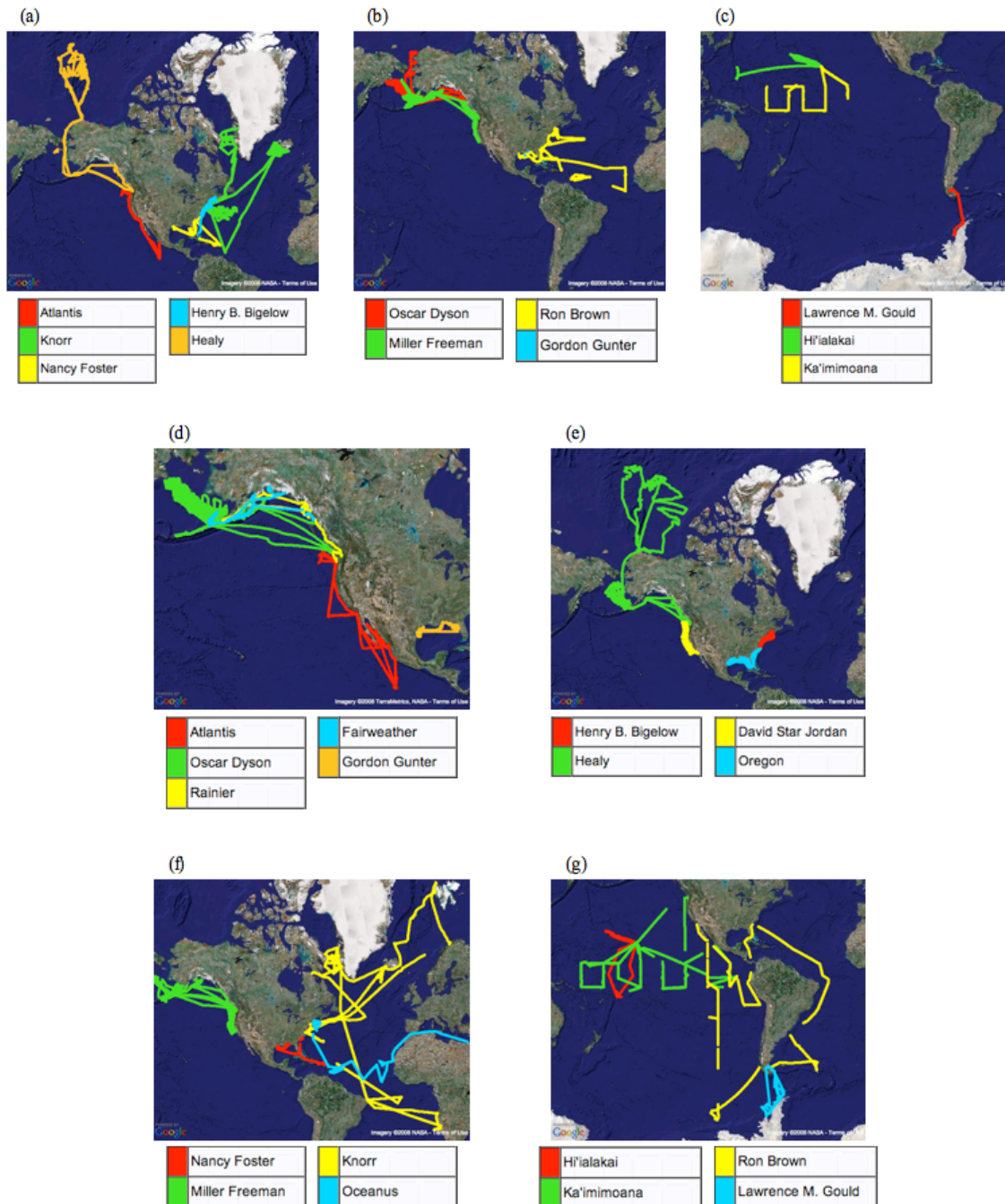
3. FIGURES AND IMAGES

Figure 1. Operational data flow between research vessels at sea and the SAMOS DAC. Data transfers take advantage of 24/7 broadband satellite communications. Real-time data quality feedback to vessels at sea and their home institutions have proven successful to reduce the amount of poor quality data caused by sensor failures.

Figure 2. Cruise tracks comparing SAMOS data coverage for FY 2007 (a, b, and c) and FY 2008 (d, e, f, and g). The maps are created using Google Earth maps through an on-line application developed by the DAC. Any user can draw cruise maps for any available SAMOS data using the tool at: http://samos.coaps.fsu.edu/html/ship_tracker.php.

4. PUBLICATIONS AND REPORTS

4.1. Technical reports

Smith, S. R., and L. Petit de la Villeon, 2008: Report of the 2nd Joint GOSUD/SAMOS Workshop. Center for Ocean-Atmospheric Prediction Studies, Florida State University, 71 pp. [Available from COAPS, The Florida State University, Tallahassee, FL 32306-2840].

4.2. Conference proceedings/presentations

Smith, S. R., and J. Rolph, 2008: SAMOS Metadata: Architecture and Lessons Learned. Presentation, *NSF Legacy of Ocean Exploration Project Data Management Meeting*, LDEO, Palisades, NY, 3-5 September 2008.

Smith, S. R., G. Goni, L. Petit de la Villeon, and M. A. Bourassa, 2008: Sustained underway oceanic and atmospheric measurements from ships: A multi-user component of the ocean observing system. Short Abstract, *NOAA Climate Observation Division 6th Annual System Review*, Silver Spring, MD, 3-5 September 2008.

Smith, S. R., J. Rolph, and M. A. Bourassa, 2008: Activities of the Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative Data Assembly Center. Short Abstract, *NOAA Climate Observation Division 6th Annual System Review*, Silver Spring, MD, 3-5 September 2008.

Smith, S. R., 2008: Obtaining quality data from an ocean observing system. Presentation, *2008 Marine Advanced Technology Education Partner Meeting*, Key West, FL, 11-13 August 2008.

Rolph, J., S. R. Smith, and M. A. Bourassa, 2008: Quality evaluation of marine meteorological observations. Short Abstract, *2nd Joint GOSUD/SAMOS Workshop*, Seattle, WA, 10-12 June 2008.

Smith, S. R., M. A. Bourassa, and J. R. Rolph, 2008: Data flow through the Shipboard Automated Meteorological and Oceanographic Systems (SAMOS) Data Assembly Center. Short Abstract, *2nd Joint GOSUD/SAMOS Workshop*, Seattle, WA, 10-12 June 2008.

Rolph, J., S. R. Smith, and M. A. Bourassa, 2008: Quality evaluation of marine meteorological observations. Short Abstract, *2008 Northern Gulf Institute Annual Conference*, Biloxi, MS, 13-14 May 2008.

Smith, S. R., M. A. Bourassa, and J. R. Rolph, 2008: The Shipboard Automated Meteorological and Oceanographic Systems (SAMOS) Initiative. Short Abstract, *Third JCOMM Workshop on Advances in Marine Climatology (CLIMAR-III)*, Gdynia, Poland, 6-9 May 2008.

Smith, S. R., M. A. Bourassa, S. D. Woodruff, S. J. Worley, E. C. Kent, and N. Rayner, 2008: A project to create bias-corrected marine climate observations from ICOADS. Short Abstract, *Third JCOMM Workshop on Advances in Marine Climatology (CLIMAR-III)*, Gdynia, Poland, 6-9 May 2008.

Woodruff, S. D., P. Brohan, E. C. Kent, S. J. Lubker, R. W. Reynolds, S. R. Smith, and S. J. Worley, 2008: ICOADS: Data Characteristics and Future Directions. Short Abstract, *Third JCOMM Workshop on Advances in Marine Climatology (CLIMAR-III)*, Gdynia, Poland, 6-9 May 2008.

Smith, S. R., J. Rolph, and M. A. Bourassa, 2008: The Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative. Short Abstract, *2008 Ocean Sciences Meeting*, Orlando, FL. 2-7 March 2008.

Smith, S. R., 2007: Metadata automation: Survey Results and Ideas. Presentation, *UNOLS Research Vessel Technical Enhancement Committee 2007 Annual Meeting*, Moss Landing, CA, 6-8 November 2007.

Smith, S. R., and J. Rolph, 2007: The SAMOS Initiative. Presentation, *UNOLS Research Vessel Technical Enhancement Committee 2007 Annual Meeting*, Moss Landing, CA, 6-8 November 2007.